

Hot-Spot Identification

Categorical Binary Model Approach

Sara Ferreira and António Couto

An alternative methodology is presented for hot-spot identification based on a probabilistic model. In this method, the ranking criterion for hot-spot identification conveys the probability of a site's being a hot spot or not being a hot spot. A binary choice model is used to link the outcome to a set of factors that characterize the risk of the sites under analysis on the basis of two categories (0/1) for the dependent variable. The proposed methodology consists of two main steps. After a threshold value for the number of accidents is set to distinguish hot spots from safe sites (Category 1 or 0, respectively), a binary model based on this classification is applied. This model allows the construction of a site list ordered by using the probability of a site's being a hot spot. In the second step, the selection strategy can target a fixed number of sites with the greatest probability or all sites exceeding a specific probability, such as .5. To demonstrate the proposed methodology, simulated urban intersection data from Porto, Portugal, covering 5 years are used. The results of the binary model show a good fit. To evaluate and compare the probabilistic method with other commonly used methods, the performance of each method is tested by its power to detect true hot spots. The test results indicate the superiority of the proposed method. This method is simple to apply, and critical issues such as assumptions of a prior distribution effect and the regression-to-the-mean phenomenon are overcome. Further, the model provides a realistic and intuitive perspective.

The identification and correction of hot spots (also referred to as black spots, hazard sites, high-risk sites, accident-prone sites, sites with promise, or priority investigation locations) is a key avenue by which traffic engineering aims to reduce the occurrence of future accidents. To that end, the *Highway Safety Manual* has dedicated a chapter to the network screening process (1). It includes several performance measures, from average crash frequency (AF), which is the least complex, to the excess expected average crash frequency with empirical Bayesian (EB) adjustments, which is the most complex. Each measure is described in the manual with respect to three main characteristics as the key criteria for selecting network screening performance measures: data availability, regression-to-the-mean bias, and method of establishing the performance threshold.

Some of the network screening performance measures described in the *Highway Safety Manual* are based on the EB method to account for the regression-to-the-mean phenomenon. However, some limitations of the EB method have led to the application of the full Bayesian (FB) method (2–4). These limitations include

the requirement for a large sample of data to develop safety performance functions, the lack of flexibility in defining underlying distributions for the observed accidents, and the absence of any penalty for an overparameterized model. Bayesian analysis combines information from the accident data with prior knowledge about it in a posterior distribution (3). The main difference between the EB and FB approaches is the way in which the parameters from prior distributions (hyperparameters) are determined (3). In the FB method, hyperparameters are determined on the basis of some prior belief about the behavior of the data involved (5, 6). However, specifying prior beliefs about data is challenging and controversial, and for that reason researchers have used the EB approach (3, 5). Although this approach is simpler to apply, it has been criticized for implicitly using the data twice. That is, the data are first used to estimate the parameters of the prior distribution by a maximum likelihood technique, and once these values are determined, the accident history of each site is used to make inferences about the posterior distribution (3, 5). Alternative statistical models using the Bayesian approach have been studied in relation to the prior distribution and to the effect of allowing variability in the dispersion parameter, which can lead to considerably different lists of hazardous locations, as noted by Miranda-Moreno et al. (3). In sum, although Bayesian analysis provides better results in terms of hot-spot identification, as noted by several authors (4, 7–9), the method requires some previous efforts, and it is therefore sometimes not feasible as a tool to support safety decisions.

As stated by Cheng and Washington, it is extremely difficult to identify clear differences between safe and unsafe sites because the distinction between the two may appear wholly arbitrary (10). Building on this idea, a probabilistic method for identifying hot spots based on a qualitative response model is proposed. This approach is a simple alternative that retains some important assumptions and may overcome some limitations of the most commonly used methodologies.

In the methodology proposed, hot-spot identification relies on the probability of a site's being a hot spot by considering the effects of the risk factors that characterize the site. Therefore, the ranking criterion (also called the decision parameter) is the probability associated with a site for taking on one of two possible designations: a safe site or a hot spot. A general framework of probability models is used to link the outcome to a set of factors. A binary choice model was selected because there are two categories (0/1) for the dependent variable. The proposed methodology was applied to intersections, and thus the risk factors are related to the main characteristics usually considered when reference populations are established for intersections (1): traffic volume (minor or major approaches), number of legs (three or four), and type of signalization (signalized or unsignalized). The model estimates the probability associated with both categories on the basis of these risk factors.

Faculty of Engineering, Civil Engineering Department, Porto University, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal. Corresponding author: S. Ferreira, sara@fe.up.pt.

Transportation Research Record: Journal of the Transportation Research Board, No. 2386, Transportation Research Board of the National Academies, Washington, D.C., 2013, pp. 1–6.
DOI: 10.3141/2386-01